

# Revisiones Sistemáticas: Recomendaciones para un Proceso Adecuado a la Ingeniería del Software

Oscar Dieste<sup>1†</sup>, Anna C. Grimán<sup>2</sup>, Natalia Juristo<sup>1</sup> y Marta López<sup>3</sup>

<sup>1</sup> Departamento de Lenguajes y Sistemas Informáticos e Ingeniería de Software  
Universidad Politécnica de Madrid  
Campus de Montegancedo. 28660. Boadilla del Monte. Madrid. España  
odieste@fi.upm.es

<sup>2</sup> Departamento de Procesos y Sistemas  
Universidad Simón Bolívar, Edf. Matemáticas y Sistemas. Caracas. Venezuela.  
agriman@usb.ve

<sup>3</sup> Departamento de Ingeniería del Software e Inteligencia Artificial.  
Universidad Complutense de Madrid.  
Ciudad Universitaria. 28040. España  
mlf@fdi.ucm.es

**Resumen.** En la actualidad, se han realizado varias revisiones sistemáticas en Ingeniería del Software y la tendencia señala que su número irá en aumento en el futuro inmediato. La práctica totalidad de las revisiones se realizan siguiendo alguna de las recomendaciones existentes, tales como las propuestas por Kitchenham, Biolchini y otros. Sin embargo, dichas recomendaciones pasan por alto una serie de carencias específicas de la experimentación en Ingeniería del Software, tales como carencias en la formación de los revisores o disponibilidad de información detallada acerca de los estudios primarios, que inducen problemas a la hora de realizar la revisión. Como resultado de analizar nuestro desempeño en la realización de varias revisiones, los autores hemos identificado diversas disfunciones. Tras su catalogación y análisis se han derivado recomendaciones para mejorar los procedimientos de revisión sistemática en Ingeniería del Software.

**Palabras Clave:** Revisión sistemática, Ingeniería del Software, Recomendaciones.

## 1 Introducción

En los últimos años, se han publicado varias revisiones sistemáticas (RS) en Ingeniería del Software (IS), por ejemplo [4][8][9][10][12][14][15][17][19][22][23][25][26][27][29]. Dichas RS persiguen aprovechar los hasta ahora dispersos resultados experimentales para obtener fragmentos de conocimiento que permitan mejorar el desarrollo de software [11]. Siguiendo esta misma tendencia, los autores también hemos realizado varios trabajos de revisión, tales como [16] [4] (y algunos otros trabajos en curso de realización), así como una actualización de las conclusiones alcanzadas en una RS previa mediante la adición de nuevas evidencias [7].

En este artículo no se pretende exponer los resultados alcanzados, los cuales fueron ya publicados en otros foros, sino resaltar ciertos problemas inesperados que incrementaron el esfuerzo para realizar dichas RS por encima de lo esperado, llegando incluso en ocasiones a amenazar el éxito de las mismas. Por ejemplo, algo tan simple como comparar las técnicas de educación registradas en [4] con las obtenidas en los nuevos estudios se demostró una tarea extremadamente compleja, al no haber tomado en consideración ni disponer de una breve

---

<sup>†</sup> Actualmente en Fraunhofer IESE, Fraunhofer Platz 1, 67663, Kaiserslautern, Alemania.  
Email: oscar.dieste@iese.fraunhofer.de

descripción de las técnicas. No es un problema trivial si se considera que el conocimiento que soporta cualquier práctica basada en evidencias no es estático. Por el contrario, es fundamental que los resultados obtenidos de una RS sean mantenidos en el tiempo al incorporar nuevas y confiables evidencias.

Cómo este, se han localizado otros problemas que afectan a la práctica totalidad de las actividades de RS. A nuestro juicio, dichos problemas no han sido adecuadamente considerados en los procedimientos de RS habitualmente empleados en IS, como los descritos en [1][20][21], siendo necesario tomar acciones para limitar su impacto. Señalar qué aspectos es necesario corregir es el objetivo del presente artículo. Para ello, se empleará la siguiente estructura: en la sección 2 se detallan los escasos procesos de RS en IS. La sección 3 incluye un esbozo de trabajos relacionados, respecto a críticas y mejoras de procesos de RS en IS. La sección 4 identifica los problemas detectados en las RS que hemos llevado a cabo. Todos estos resultados se discuten en la sección 5 y finalmente se presentan las conclusiones en la sección 6.

## 2 Antecedentes

Existen pocas iniciativas en cuanto a procesos de RS en IS. La propuesta pionera en esta área es la presentada por [20] que ha sido aplicada, con pocas modificaciones, en todos los procedimientos de RS identificados en la literatura. Es evidente que el procedimiento propuesto por [20] recibió una influencia directa de los procedimientos para RS utilizados en medicina, muy especialmente el propuesto por [18]. Las tres fases del procedimiento de Kitchenham son las siguientes:

- *Planificación de la revisión.* Su objetivo es determinar si hay una verdadera necesidad de la revisión y desarrollar un plan para ejecutar la misma. Esta fase se divide en dos actividades:
  - Identificación de la necesidad de la revisión: su objetivo es evitar la duplicidad para que cada nueva revisión sea original. Si existen revisiones previas debería evaluarse su calidad con base en su objetivo, fuentes de búsqueda, criterios de inclusión/exclusión, criterios de evaluación de la calidad, métodos de extracción y síntesis de datos.
  - Desarrollo de un protocolo de revisión: este protocolo especifica los métodos a ser utilizados en la RS. Se pretende evitar los sesgos producidos por cambios en la *pregunta de investigación* o por una selección subjetiva de los estudios primarios. Este protocolo contiene: los antecedentes de la revisión, las preguntas de investigación, la estrategia de búsqueda, los criterios y procedimientos para la selección de estudios y para la evaluación de su calidad, la estrategia de extracción de datos y los métodos para su síntesis.
- *Conducción de la revisión.* Esta fase presenta actividades muy similares a las propuestas por [18] y [13]. Consta de las siguientes cinco actividades:
  - Identificación de la investigación: consiste en la identificación de estudios primarios relevantes a través de una estrategia de búsqueda que debe ser exhaustiva y libre de sesgo.
  - Selección de los estudios primarios: consiste en determinar la relevancia de cada estudio identificado, respecto a la pregunta de investigación planteada. Esta tarea se realiza con base en los criterios de inclusión/exclusión establecidos en el protocolo de revisión.
  - Evaluación de la calidad de los estudios: su propósito es obtener información adicional acerca de los estudios seleccionados para refinar los criterios de inclusión/exclusión. En este contexto, la calidad es definida en relación a la manera en que cada estudio reduce la probabilidad de sesgos y maximiza la validez interna y externa de sus resultados. Para [20] la evaluación de la calidad se enfoca en asignar un peso a cada estudio y establecer si las diferencias en cuanto a la calidad de un estudio respecto a otro pueden explicar resultados diferentes entre ellos. Aunque [20] reconoce la necesidad de incorporar estudios de tipos o diseños diversos, propone en esta actividad como criterio principal para la determinación de la calidad, una jerarquía de diseños para la IS donde los estudios del tipo *experimento controlado* ocupan el primer nivel (tal como ocurre en jerarquías similares propuestas para

medicina). También propone algunos criterios genéricos (relacionados con el diseño) y específicos (relacionados con el tópico de revisión) tomados de [13].

- Extracción de datos y monitoreo: el objetivo es diseñar formularios para realizar la recolección de datos a partir de los estudios primarios de manera precisa y libre de sesgo.
- Síntesis de datos: el propósito es resumir y agregar los datos que han sido extraídos de los estudios primarios seleccionados. Puede realizarse a través de la tabulación o descripción de las características de los estudios y sus resultados (síntesis no-cuantitativa) o de una síntesis cuantitativa (meta-análisis) a través de la aplicación de técnicas estadísticas. En [20] se plantea que los procedimientos a ser utilizados deben ser especificados anticipadamente en el protocolo de revisión.
- *Reporte de la revisión.* Como última fase de la RS, se debe generar un reporte que permita a los investigadores y profesionales entender las implicaciones de los resultados de la RS, al mismo tiempo que pueda valorar su validez. Para ello [20] sugiere comunicar los resultados a través de diferentes medios: reportes técnicos, una sección en una tesis doctoral, artículos técnicos en revistas y conferencias, artículos no técnicos en revistas para profesionales, en notas de prensa, y en páginas Web. El procedimiento propone una estructura guiada por el protocolo de investigación donde se documenta información resumida de los componentes del protocolo y una discusión breve de resultados y conclusiones.

La propuesta original de [20] ha sido modificada por Biolchini y sus colegas [1] quienes proponen un proceso que comienza con la *planificación de la revisión*, donde se genera un protocolo de revisión que debe ser aprobado. Posteriormente, se ejecuta la revisión (lo cual incluye la *búsqueda, selección y evaluación de estudios*, así como la *extracción de los datos*) y, finalmente, se analizan los datos obtenidos a través de métodos de *meta-análisis*. De manera concurrente, durante todo el proceso, los resultados que se obtienen son “empaquetados” en una base de datos.

También en esta propuesta los autores proponen que algunas de las fases sean iterativas; esto es, las actividades son iniciadas durante el desarrollo del protocolo y revisadas mientras se conduce la RS (p.e. la definición de los criterios de inclusión/exclusión, la extracción de datos, y la síntesis de datos). Sin embargo, su principal aporte es la propuesta de una plantilla para el desarrollo del protocolo de revisión. Esta plantilla establece, aunque sin mucho detalle, la información que debe ser proporcionada en la fase temprana del proceso (formulación de la pregunta, selección de fuentes, selección de estudios, extracción de información, y resumen de resultados).

Por otra parte, [4] proponen modificaciones en las fases de conducción de la revisión y en la síntesis de datos. En particular, propone técnicas de codificación de los datos extraídos de los estudios primarios así como un protocolo de agregación basado en un perfeccionamiento del análisis comparativo que permite la agregación sistemática de los datos experimentales obtenidos. Este protocolo de agregación se soporta en una notación que permite la manipulación puramente formal de los datos, lo que evita que el revisor introduzca sesgos originados en sus preferencias o por juicios subjetivos acerca de los tratamientos involucrados en los estudios primarios. Los autores también aportan una estrategia de generalización de los tratamientos y variables respuesta que permite sobrellevar el problema de la falta de uniformidad y obtener resultados generalizados de grano grueso que son aptos para la agregación.

### 3 Trabajos relacionados

De los antecedentes que hemos señalado previamente, la propuesta de Kitchenham es sin duda la más extendida y ampliamente utilizada. Es por ello que la mayoría de las RS que encontramos en la literatura ([4][8][9][10][12][14][15][17][19][22][23][25][26][27][29]) han aplicado, más o menos fielmente, los lineamientos propuestos por Kitchenham [20]. Vemos entonces que las experiencias recogidas por los revisores durante los últimos años han dado lugar a la confirmación

de algunos de los lineamientos propuestos en [20] pero también a críticas y propuestas de mejora. Algunas de ellas son descritas a continuación relacionadas con las etapas del proceso de RS.

Una de las primeras propuestas de cambio fue introducida en [2] y [3] donde se reconoce la dificultad de establecer un protocolo de investigación definitivo en una etapa temprana del proceso de RS y se recomienda la realización de un estudio piloto del protocolo. En estos trabajos también se reconoce que las preguntas de investigación deberían ser revisadas durante el desarrollo del protocolo, a medida que los revisores van ganando entendimiento del problema. Esto implica que los autores reconocen la dificultad de realizar el proceso de revisión de manera absolutamente lineal como pareciera sugerir la propuesta original de Kitchenham.

Estas ideas son retomadas en [28] quienes consideran el desarrollo del protocolo de investigación como un punto crítico. A pesar de que reconocen la utilidad del estudio piloto, los autores señalan que los lineamientos proporcionados no dan una orientación específica acerca de los criterios de parada del mismo, lo cual puede ser la causa de sesgos y errores futuros. Abordan también la necesidad de establecer lineamientos acerca del reporte de las modificaciones sufridas por el protocolo de investigación original; en este sentido indican que no existe claridad acerca de si se debe reportar (solamente) la versión final del protocolo o si se debe documentar también el protocolo inicial junto con las razones que justifican los cambios sufridos por éste.

Por otra parte, en [28] también se señalan otros aspectos neurálgicos del proceso de RS de Kitchenham. Uno de ellos es el gran esfuerzo requerido para formular las preguntas de investigación y definir la unidad de análisis para la RS. Finalmente, abordan también otras etapas del proceso; es así como destacan la necesidad de establecer lineamientos para: 1) evaluar de manera válida y confiable diferentes tipos de estudios no contemplados en [20], y 2) realizar chequeos de confiabilidad entre resultados de los diferentes revisores o evaluadores. Por último, aunque no relacionado con la propuesta de Kitchenham, los autores señalan la necesidad de crear y mantener un índice centralizado de RS en IS.

Otro de los estudios que ha abordado la problemática alrededor de la aplicación de los lineamientos de Kitchenham para las RS en IS es [8]. Luego de su aplicación a un caso práctico, estos autores enfocan sus recomendaciones principalmente en las actividades de evaluación de la calidad de los estudios y el análisis de los resultados. Algunas de sus recomendaciones para mejorar la propuesta de [20] son las siguientes:

- Incluir evidencias desde una variedad de perspectivas y métodos de investigación.
- Proporcionar mayores lineamientos acerca de la conducción de RS utilizando enfoques cualitativos, como la meta-etnografía, y métodos mixtos; y utilizar las técnicas y terminología usualmente asociada con el análisis cualitativo de datos de investigación primaria.
- Proporcionar métodos específicos para la evaluación de la calidad, p.e. enfoques estructurados, y la síntesis de los datos empíricos extraídos de los resultados de los diversos tipos de estudio.
- Explorar el uso de software para soportar los métodos de síntesis cualitativa, y
- Desarrollar mejores maneras de integrar la síntesis cualitativa con el meta-análisis.

Finalmente, en [2] se exponen algunas lecciones aprendidas luego de la experiencia de aplicar el proceso propuesto en [20] a tres revisiones. A continuación describimos aquellas que no han sido mencionadas previamente por otros autores:

- Realizar un estudio de mapeo previo a la revisión para ayudar a establecer las preguntas de investigación.
- Participación de todos los miembros del equipo en el desarrollo del protocolo.
- Extracción de datos asistida por: 1) las definiciones de los datos y, 2) lineamientos específicos a la extracción, en documentos separados.
- Un proceso de validación del protocolo realizado por revisores externos.
- Seleccionar y justificar una estrategia de búsqueda adecuada a la pregunta de investigación, asimismo realizar la búsqueda en diferentes fuentes electrónicas.

- Debido a la pobre calidad de los resúmenes de las publicaciones en IS, deberían revisarse no sólo éstos sino también las conclusiones del estudio antes de seleccionar aquellos estudios primarios que participarán de la RS.
- Asegurarse de cómo usar la calidad en la posterior agregación y análisis de datos para establecer si la evaluación de la calidad es necesaria así como los métodos de evaluación adecuados. Una aproximación es presentada en [19] donde se plantea la evaluación en dos dimensiones: 1) la calidad del estudio en sí mismo, y 2) la calidad del reporte proporcionado.
- Utilizar dos roles para la extracción de datos cuando se cuente con una gran cantidad de estudios: uno que realice la extracción y otro que realice la verificación.
- Utilizarse la tabulación de datos en lugar de meta-análisis en IS, siempre que los revisores se aseguren de explicar cómo los datos agregados responden las preguntas de investigación.
- Mantener un registro de las decisiones tomadas durante el proceso de revisión.

Se observa que las recomendaciones aportadas por el resto de los autores se enfocan más en los aspectos más críticos del proceso de revisión, pero las propuestas en [2] buscan abarcar todas las etapas del proceso. No obstante, el nivel de especificidad de la mayoría de las recomendaciones no es suficiente. Por ejemplo, la recomendación “seleccionar y justificar una estrategia de búsqueda adecuada a la pregunta de investigación”, es demasiado vaga para ser aplicada directamente en la práctica. Lo mismo ocurre con otras recomendaciones como “incluir evidencias desde una variedad de perspectivas y métodos de investigación” la cual, por su generalidad, no aporta una solución concreta a la carencia de orientaciones metodológicas para la conducción de RS.

Éstas y otras recomendaciones generales dejan un amplio margen de discrecionalidad al revisor, lo cual es deseable para evitar rigidez en la aplicación del proceso. En contrapartida, se abstraen en exceso del estado de la experimentación en IS, ocultando la complejidad real de las tareas a realizar y no anticipando problemas y disfunciones que deben ser resueltas específicamente por los revisores, obligándoles a “reinventar la rueda” en cada nueva RS.

## 4 Problemas identificados

Nuestra experiencia en la realización de RS nos ha permitido aprender un conjunto de lecciones derivadas de los problemas encontrados tanto durante la ejecución inicial de una RS [16][4] como en su posterior actualización [7]. Esto nos ha permitido desarrollar una visión de las fortalezas y debilidades de los procesos en uso actualmente.

Un ejemplo ayudará a aclarar nuestra postura. Por ejemplo, cuando se intenta aplicar meta-análisis a los datos extraídos de los estudios primarios. En general, la aplicación de esta técnica estadística es muy limitada, ya que los datos son siempre escasos y a menudo incompatibles. Este problema ya ha sido señalado por otros autores como [9], que propone complementar las técnicas de síntesis cuantitativas con otras técnicas cualitativas.

Al igual que este problema, existen un buen número de otros problemas de carácter menos metodológico y más ligados a la práctica que no han sido mencionados en la literatura y que afectan a la práctica totalidad de las actividades de RS. Dichos problemas, así como las actividades a las que afectan, se muestran en la Tabla 1 (confeccionada utilizando la terminología de Kitchenham [20]), y se describen en las secciones siguientes.

### 4.1. Cualificación de los revisores

Una primera serie de problemas tiene relación con la cualificación de los revisores. Hemos observado que el desarrollo de una RS requiere de *conocimientos específicos acerca de experimentación*. Sino pueden producirse disfunciones como las siguientes:

- Desarrollo del protocolo de la revisión: definición deficiente de la pregunta de investigación, definición de criterios de inclusión/exclusión demasiado generales.

**Tabla 1.** Resumen de problemas encontrados y relación con las actividades de una RS.

<div> <div>PROBLEMAS DETECTADOS</div> <div>ACTIVIDADES DE LA REVISIÓN</div> </div>		Cualificación de los revisores		Contenido de las plantillas de extracción y reporte final		Cantidad y calidad de los estudios primarios		Proceso
		Falta de conocimientos de experimentación	Falta de conocimientos del área bajo revisión	Ausencia de trazabilidad a artículos originales	No disponer de un glosario de términos	Carencia de instrumentos de evaluación de la calidad	Acotar la identificación de estudios primarios	Proceso demasiado genérico
Planificación	Identificación de la necesidad de la revisión							
	Desarrollo del protocolo de revisión							
Conducción	Identificación de la investigación							
	Selección de los estudios primarios							
	Valoración de la calidad de los estudios							
	Extracción de datos							
	Síntesis de datos							
Reporte de la revisión								

- Selección de estudios: no aplicar correctamente los criterios de inclusión y exclusión, incrementar el número de artículos candidatos que finalmente no son seleccionados (p.ej., pre-seleccionando artículos que sugieren pero no realizan experiencia empírica alguna).
- Valoración de la calidad: no identificar los distintos tipos de estudios empíricos (experimentos, cuasi-experimentos, casos de estudio, etc.); ignorar las características de un estudio bien diseñado y ejecutado (aleatorización, control de los parámetros experimentales, etc.); no valorar si el experimento está o no correctamente analizado.
- Extracción de datos: no identificar los factores y niveles de un experimento multifactorial; no distinguir entre variables respuesta y constructos del investigador (secciones de discusión); no entender el significado de obtener resultados contradictorios en distintos experimentos.

El desarrollo de una RS requiere, además de conocimientos generales de experimentación, *conocimientos específicos acerca del área de revisión*. Sino, es extremadamente difícil:

- Decidir si las RS existentes son adecuadas en función del estado del arte del área considerada.
- Identificar correctamente los estudios empíricos relevantes.
- Identificar las similitudes existentes entre distintos estudios empíricos compatibles (cuando dos experimentos ensayan el mismo tratamiento o variable respuesta), lo cual dificulta la posterior síntesis.
- Identificar posibles generalizaciones de los tratamientos ensayados en diferentes estudios (en función de las definiciones proporcionadas por los experimentadores o de los conocimientos previos del revisor) que permitan realizar agregaciones válidas durante la síntesis de datos.
- Asimismo, el desconocimiento del área produce que el reporte de la RS no esté ajustado a las necesidades de los investigadores y profesionales (p.ej., se ahonda en técnicas poco relevantes o escasamente usadas porque el número de experimentos acerca de esas técnicas es elevado).

#### 4.2. Contenido de las plantillas de extracción y reporte final

Un segundo tipo de problemas tiene relación con las plantillas de extracción de datos y reporte de los resultados. Habitualmente, este tipo de materiales es bastante sumario, conteniendo sólo el mínimo de información para realizar la revisión y comunicar las conclusiones obtenidas. Sin

embargo, la falta de uniformidad terminológica (así como también de fondo) en IS provoca que tanto durante la primera realización de la RS, como durante su actualización, sea necesario volver continuamente a los estudios primarios para obtener información ausente o confirmar suposiciones. Esto provoca que la extracción de datos sea mucho más costosa y, en determinadas circunstancias, puede también perjudicar la síntesis de datos.

Un caso típico ocurre cuando un artículo hace referencia a una técnica genérica (p.e. un tipo determinado de entrevista o de inspección). ¿Cómo es posible decidir si esta técnica es similar a alguna considerada con anterioridad, de modo que los estudios respectivos pudieran agregarse? Habitualmente, las plantillas de extracción y/o los reportes de RS previas no contienen la información necesaria, por lo que es necesario releer los estudios primarios previamente revisados. Lo deseable sería disponer de *enlaces entre las plantillas de extracción y los estudios primarios*, de modo que fuese posible consultar directamente la información utilizada durante la extracción de datos, en lugar de buscar dicha información de nuevo por todo el artículo. Asimismo, disponer de un *glosario de términos* (que incluyese como mínimo a los factores y variables respuesta, aunque podría ampliarse a otros aspectos de los artículos, tales como los parámetros experimentales) que resumiese las decisiones tomadas por los revisores con anterioridad (p.e. por qué han considerado que la técnica X e Y son similares) sería igualmente útil.

### 4.3. Cantidad y calidad de los estudios primarios

Un tercer tipo de problemas tiene relación con la relación entre la cantidad/calidad de los estudios y el número de evidencias que éstos proporcionan. La figura 1 muestra la relación encontrada en [4] entre el número de estudios primarios pre-seleccionados y la cantidad de evidencias obtenidas. Tal y como puede apreciarse, la cantidad de evidencias localizadas es considerable durante los primeros estadios del proceso de identificación de estudios primarios, pero disminuye estrepitosamente a medida que éste avanza. El incremento final se debe a 2 publicaciones identificadas al final del proceso que deben ser consideradas como outliers para este análisis.

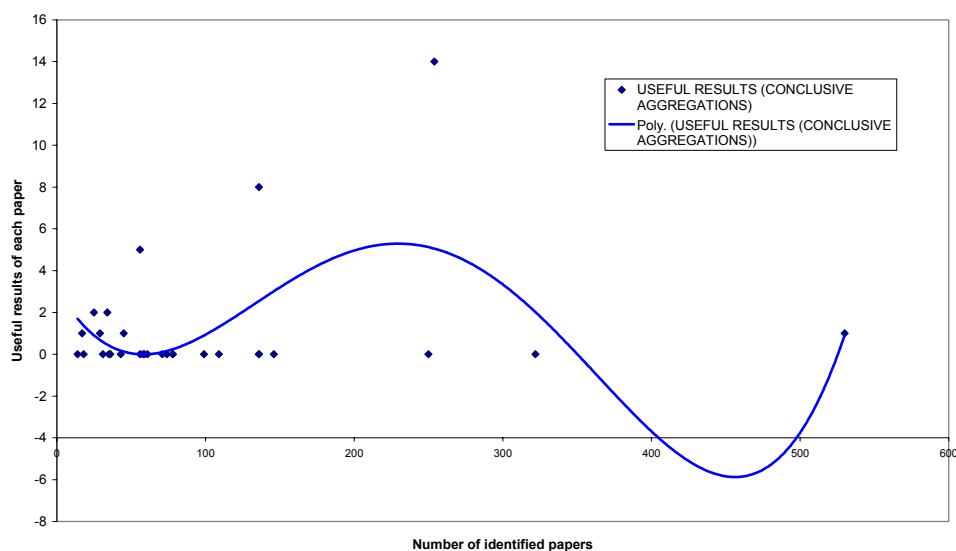
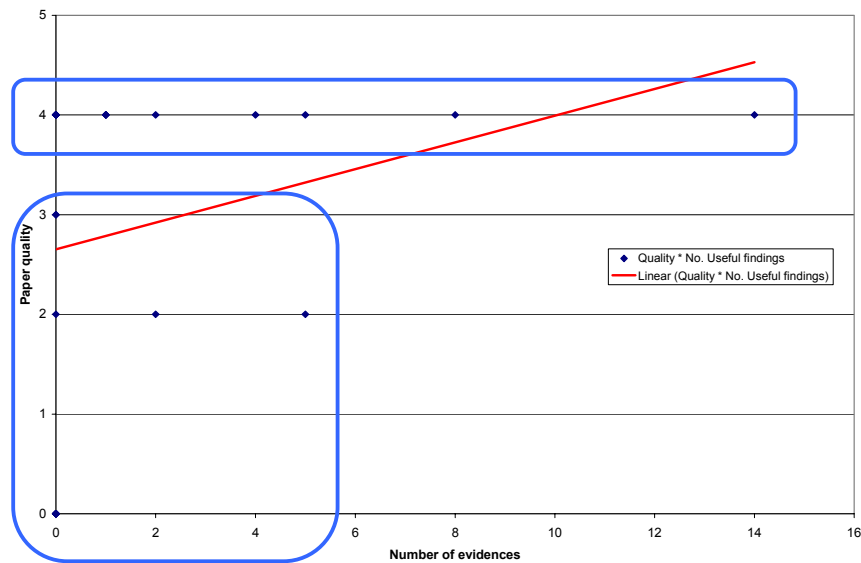


Fig. 1. Líneas de tendencia respecto al número de pre-seleccionados.

Los datos de la figura 1 sólo permiten concluir que el *proceso de identificación de estudios primarios debe estar acotado* de algún modo que permita evitar localizar una cantidad elevada de estudios primarios inútiles que deben ser filtrados durante el proceso de selección, con el esfuerzo que ello supone. Dicho de otro modo: los revisores deben intentar que los procesos de

identificación y selección sean costo-efectivo utilizando para ello estrategias de búsqueda sistemáticas tales como las propuestas en [6], que: 1) maximicen la cantidad de material relevante y 2) economicen el esfuerzo de descarte.

En línea con lo anterior, la figura 2 muestra la relación entre la calidad de los estudios primarios y las evidencias de ellos obtenidas. La calidad de los estudios se ha valorado utilizando un instrumento propio de [4], ya que no existen instrumentos normalizados de valoración de la calidad. Este instrumento asigna a cada estudio un valor en función de, por un lado, el tipo de diseño experimental utilizado (de un modo similar a lo propuesto por Kitchenham [20]), y por otro al grado en que el experimento ha sido correctamente diseñado y ejecutado.



**Fig. 2.** Relación entre la calidad de los artículos y los resultados obtenidos.

Como se observa, una alta calidad no va asociada a un alto número de evidencias, pero una baja calidad decididamente implica un bajo número de evidencias. Es decir, una *evaluación temprana de la calidad* es vital para reducir el esfuerzo de revisión, sobre todo en el número de artículos que es necesario leer durante la extracción de datos. Pero al no existir instrumentos normalizados de evaluación, todo queda al arbitrio del revisor, con el riesgo de que la evaluación sea demasiado (o poco) exigente y que la cantidad de evidencias sea escasa (o el esfuerzo de extracción muy alto).

#### 4.4. Proceso

Finalmente, también se debe resaltar el problema del grado de abstracción que, en líneas generales, poseen todas las recomendaciones acerca de cómo realizar RS en IS. Ello contrasta netamente con el detalle de las recomendaciones en otras disciplinas, como es el caso de [13]. Un proceso muy genérico deja demasiados aspectos al arbitrio de los revisores, complicando desde nuestro punto de vista la realización de actividades de valoración de calidad, extracción de datos y síntesis de datos.

## 5 Discusión

La escasa cantidad de procesos de RS en IS, y su uso casi estandarizado, podría inducir a pensar que dichos procesos ([1][20][21]) están lo suficientemente detallados como para que no surjan



problemas a la hora de llevar a cabo una RS. Sin embargo, como se ha mostrado en las secciones anteriores, sí se detectan diversos problemas. Aunando todas las fuentes consideradas, hemos obtenido las recomendaciones de mejora de los procesos de RS que se detallan a continuación.

- Respecto al rol de revisor, definir las cualificaciones mínimas que debe poseer un revisor. No sólo se debe pensar en los conocimientos mínimos que un revisor debe tener en cuanto al área a revisar sino también del propio proceso de revisión, que implica poseer conocimientos básicos sobre experimentación.

Alternativamente, se sugiere definir un conjunto de roles con perfiles claramente identificados para la realización efectiva de una RS. Al igual que ocurre en procesos de desarrollo de software, por ejemplo, si se cuenta con una relación de roles se facilita mucho su asignación a las personas adecuadas, respecto a su capacidad, conocimientos y experiencia. En tal sentido, proponemos el siguiente conjunto mínimo de roles a incluir en el proceso de RS:

- Investigador principal: encargado de promover la RS, establecer el tema sobre el cual se realizará ésta, supervisar y controlar el proceso. Debe poseer amplios conocimientos de experimentación y experiencia en gestión de proyectos de investigación. Es responsable de la definición del objetivo de revisión y los criterios de inclusión/exclusión de estudios.
  - Analista de búsquedas: encargado de la localización de los estudios empíricos en las bases de datos bibliográficas y otras fuentes (p.e. bibliografías, Internet, etc.), el desarrollo de y documentación de estrategias de búsqueda óptimas, incluyendo la selección de repositorios, campos de búsqueda, cadenas de búsqueda y el universo de búsqueda, y de mantener el control de las referencias utilizadas durante la RS. Debe poseer conocimientos específicos sobre bases de datos bibliográficas y publicaciones referidos a la IS.
  - Analista de tópico: un experto que realiza funciones de soporte al Analista de Búsqueda, específicamente en el desarrollo de la cadena de búsqueda adecuada al tema de revisión seleccionado. Debe poseer conocimientos avanzados en el área de IS o manejo de la literatura especializada.
  - Analista de datos experimentales: una vez seleccionados los estudios que participarán en la RS, los datos experimentales deben ser extraídos de los mismos y presentados de una manera estándar. Estas funciones son realizadas por un rol con conocimiento de IS y en los posibles factores ensayados en los estudios primarios.
  - Evaluador de calidad: este rol es el encargado de valorar la calidad de cada estudio seleccionado, así como de clasificar los estudios de acuerdo con su nivel de fiabilidad. Debe poseer conocimientos en el método de valoración de la calidad a aplicar y los criterios inherentes al mismo.
- Contar con pautas para el establecimiento de la pregunta de investigación y la descripción del protocolo de revisión. Se debe ser consciente de que: 1) es preciso llegar a un punto de equilibrio en la generalidad-especificidad que se use a la hora de plasmar la pregunta; y 2) el protocolo se modificará a lo largo de la RS ineludiblemente. Para ello, pensamos que la realización temprana de un *estudio piloto*, a través de la búsqueda preliminar de estudios y la clasificación de éstos con base al tópico estudiado y el diseño experimental, permitirá establecer tanto la pregunta de investigación como los métodos a utilizar para el análisis y la síntesis de datos. La exhaustividad con la que se realice este estudio preliminar determinará el nivel de especificidad con el que se plantee la pregunta de investigación y el protocolo.
  - Utilizar estrategias de búsqueda sistemáticas para la identificación y selección de estudios primarios de interés. Se pretende que la RS sea también un proceso costo-efectivo, especialmente en estas actividades iniciales. Afortunadamente, a este respecto ya existen trabajos tales como [6], así como otras investigaciones en curso. En tales estudios se ha comprobado la costo-efectividad del uso de cadenas de búsqueda óptimas que maximizan la recuperación de artículos relevantes. Ejemplo de una cadena que, en nuestra experiencia, puede proporcionar resultados óptimos para la recuperación de experimentos controlados es la combinación de los términos *experiment*, *experimental study*, *experimental comparison*, *experimental analysis*, *experimental evidence*, y *experimental setting*.

- Aumentar el detalle de los procesos de RS. Como se ha mencionado a lo largo del artículo, hay diversas tareas concretas que precisan ser descritas con más detalle pero de entre todas ellas queremos destacar:
  - La valoración de la calidad: debería integrar criterios para la evaluación de experimentos controlados, así como otros tipos de estudios, en el contexto de la IS. Esto debería incluir propuestas obtenidas de diferentes disciplinas que puedan ser aplicables a nuestra área. Algunas propuestas que deben ser consideradas son presentadas en la Tabla 2.

**Tabla 2.** Resumen de marcos de referencia para la evaluación crítica de la calidad de los estudios empíricos [24] [5].

Escala	Tipo de estudio
Jadad	Experimentos controlados
Cowley	Estudios comparativos
Downs y Black	Estudios aleatorizados y no aleatorizados
Newcastle-Ottawa	Estudios de cohorte
Thomas, Reisch et al., Zaza et al.	Cualquier tipo de estudio
Manual de CRD	Estudios observacionales
Reporte NHS CRD	Estudios de control de caso
Grupo EPOC	Serie de tiempo interrumpidas
Egan et al.	Estudios longitudinales
Crombie	Encuestas
Spencer et al.	Investigación cualitativa

- La extracción de datos: debería definirse un conjunto de datos mínimos a ser extraídos en función de su utilidad durante la síntesis, así como la manera en que tales datos deben ser recolectados y analizados a fin de que el proceso de extracción se concentre en el grupo de estudios más prometedores para efectos de la síntesis de evidencias. En nuestra experiencia, la extracción debe ser un proceso iterativo en que se recolecten los siguientes datos: tratamientos, variables respuesta, parámetros, variables de bloque, hipótesis, sujetos experimentales, ambiente, tipo de estudio, proceso experimental, métodos de medición y análisis de los resultados, amenazas, resultados experimentales y hallazgos.
- La síntesis de datos: sería necesario desarrollar nuevas técnicas de síntesis, o adaptar las técnicas actualmente existentes en otras disciplinas, para aprovechar los hallazgos experimentales en IS que no pueden ser agregados mediante técnicas estadísticas de meta-análisis. Algunos métodos utilizados en las ciencias sociales que deberían ser considerados en lugar del meta-análisis son: la Síntesis Narrativa de estudios cuantitativos, la Síntesis de la Mejor Evidencia (BES, por sus siglas en inglés), el Conteo de Votos, la Síntesis a través de Diseños y la Triangulación.
- Disponer de formularios estandarizados de meta-datos para diversos procesos, por ejemplo:
  - Para el proceso de extracción de datos, que defina los elementos que deben ser extraídos de los estudios primarios. Ello favorecería la comprensión de los formularios de extracción y una cabal valoración de la RS y verificación de resultados obtenidos.
  - Para elaborar el reporte de la SR, contando con estándares detallados para dicho informe. Por ejemplo, usando diferentes niveles de detalle, como un tipo de reporte corto para la publicación en revista y uno detallado para su publicación en bases de datos online. Dichos reportes deberían contener algún tipo de glosario de términos.
- Sería también deseable contar con una lista de chequeo final que evalúe la documentación, conducción y resultados de la RS. Esto permitiría hacer un análisis post-mortem de la RS, analizando los problemas que hayan surgido relacionados con el proceso, roles y documentación. Pero también sería importante que se garantizara que en algún documento se

han recopilado las decisiones tomadas en la RS para que en un futuro se puedan tener en cuenta y saber cómo conducir una actualización o si, por ejemplo, sería posible llevar a cabo una agregación con resultados de otras RS.

- Finalmente, respecto al reporte de la revisión, sería deseable que incluyese como adjuntos los formularios de extracción, junto con enlaces a los estudios primarios. La ventaja de disponer de esta información es que permitiría valorar fácilmente la veracidad de los resultados de la RS, permitiendo acceder de manera directa a los estudios relacionados con cada resultado, así como realizar las actualizaciones de la revisión de manera eficiente. En la actualidad, las bases de datos online proporcionan la funcionalidad precisa para implementar esta mejora y permiten que toda esta información sea fácilmente accesible por otros investigadores.

## 6 Conclusiones

Si bien la propuesta de Kitchenham [20] ha supuesto un gran avance en el campo de la RS en la IS, consideramos que es preciso ahondar más en la investigación del proceso de RS dados los problemas detectados que afectan a la práctica totalidad de las actividades de una RS.

En este artículo se han analizado una serie de disfunciones detectadas en las RS que hemos desarrollado en el campo de la IS. El proceso que utilizamos ha sido una instanciación del propuesto por Kitchenham, como la gran mayoría de los que se llevan a la práctica en IS. Del análisis de nuestras RS se han detectado diversos problemas que han afectado a la conducción de estas revisiones y que, según nuestro juicio, no han sido adecuadamente considerados. Otros autores también han criticado los procesos típicamente utilizados en IS ([1][20][21]).

Aunando todas las críticas y puntos débiles detectados hemos obtenido una serie de recomendaciones de las que es difícil resaltar una en concreto porque prácticamente todas se pueden considerar como críticas. No obstante, podemos afirmar que las observaciones hechas al proceso propuesto por Kitchenham, así como las recomendaciones que realizamos en este trabajo, están enfocados en cinco aspectos claves del proceso que no han sido adecuadamente conducidos en las propuestas actuales, pero que afectan considerablemente el éxito de la RS y la confiabilidad de la evidencia obtenida, tales aspectos son: la sistematización de tareas, la prevención del sesgo, la presencia de heurísticas específicas a cada fase, la adecuación de las tareas a las características específicas de la base experimental en IS, y el soporte a roles.

Nuestra línea de investigación futura se orienta hacia el refinamiento de las recomendaciones aquí planteadas, así como al análisis y desarrollo de técnicas de síntesis que permitan aprovechar los hallazgos experimentales en IS que no pueden ser agregados mediante técnicas estadísticas de meta-análisis.

## Referencias

- [1] Biolchini, J., Mian, P., Natali, A., y Travassos, G. H.: COPPE/UFRJ, Rio de Janeiro, Brazil, ES 679/05, 2005.
- [2] Brereton, P., Kitchenham, B., Budgen, D., Turner, M. y Khalil, M.: Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, vol. 80, nº 4, pp: 571-583, 2007.
- [3] Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M., y Khalil, M.: *Employing Systematic Literature Review: An Experience Report*. Keele University, Department of Computer Science. 2005.
- [4] Davis, A.M., Dieste, O., Hickey, A.M., Juristo, N., y Moreno, A. M.: Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review. *Proceedings of 14th IEEE International Requirements Engineering Conference (RE'06)*, Minneapolis, USA, Sept 11-15, pp. 179-188, 2006.
- [5] Deeks, J., Dignes, J., D'Amico, R., Sowden, A., Sakarovich, C. Song, F., et al. *Evaluating non-randomized intervention studies*. Health Technology Assessment, 2003.
- [6] Dieste, O. y Griman, A. C.: Developing Search Strategies for Detecting Relevant Experiments for Systematic Reviews. *Proceedings of the 1<sup>st</sup> Int. Symposium on Empirical Software Engineering and Metrics (ESEM'02)*, 2007.

- [7] Dieste, O., Lopez, M. y Ramos, F.: Obtaining Well-Founded Practices about Elicitation Techniques by Means of an Update of a Previous Systematic Review. 20<sup>th</sup> Int. Conference on Software Engineering and Knowledge Engineering (SEKE'08), San Francisco, USA. 1-3 Julio, 2008.
- [8] Dyba, T., Arisholm, E., Sjöberg, D. I. K., Hannay, J. E., y Shull, F.: Are two heads better than one? On the effectiveness of pair programming. *IEEE Software*, vol. 24, pp. 10-13, 2007.
- [9] Dyba, T., Dingsoyr, T., y Hanssen, G. K.: Applying Systematic Reviews to Diverse Study Types: An Experience Report. 1<sup>st</sup> Int. Symposium on Empirical Software Engineering and Measurement, Madrid, Spain. 2007.
- [10] Dyba, T., Kampenes, V., y Sjöberg, D.: A systematic review of statistical power in software engineering experiments. *Information and Software Technology*, vol. 48, pp. 745-755, 2006.
- [11] Dyba, T., Kitchenham, B. A., y Jorgensen, M.: Evidence-based software engineering for practitioners. *IEEE Software*, vol. 22, n° 1, pp. 58-65, 2005.
- [12] Hannay, J. E., Sjøberg, D. I. K., y Dyba, T.: A systematic review of theory use in software engineering experiments. *IEEE Transactions on Software Engineering*, vol 33, n° 2, pp: 87-107, 2007.
- [13] Higgins, J. P. T. y Green, S.: *Cochrane handbook for systematic review of interventions*. Chichester, UK: John Wiley and Sons, Ltd., 2006.
- [14] Jørgensen, M. A.: Review of Studies on Expert Estimation of Software Development Effort. *Journal of Systems and Software*, vol 70, n° 1-2, pp: 37-60, 2004.
- [15] Jørgensen, M., y Shepperd, M.: A systematic review of software development cost estimation studies. *IEEE Transactions on Software Engineering*, vol 33, n° 1, pp: 33-53, 2007.
- [16] Juristo, N., Moreno, A. M., y Vegas, S.: Reviewing 25 Years of Testing Technique Experiments. *Journal of Empirical Software Engineering*, vol. 9, pp. 7-44, 2004.
- [17] Kampenes, V. B., Dyba, T., Hannay, J. E., y Sjøberg, D. I. K.: A systematic review of effect size in software engineering experiments. *Information and Software Technology*, vol 49, n° 11-12, pp: 1073-1086, 2007.
- [18] Khan, K., Khalid, S., Riet, G. T., Glanville, J., Sowden, A. J., y J.K.: *NHS Centre for Reviews and Dissemination. University of York, CRD Report Number 4 (2<sup>nd</sup> edition)*, 2001.
- [19] Kitchenham, B. A., Mendes, E. y Travassos, G. H.: Cross versus within-company cost estimation studies: A systematic review. *IEEE Transactions on Software Engineering*, vol. 33, n° 5, pp: 316-329. 2007.
- [20] Kitchenham, B.A.: *Procedures for performing systematic reviews*, Keele University TR/SE-0401, 2004.
- [21] Mendes, E.: A Systematic Review of Web Engineering Research. *ACM/IEEE International Symposium on Empirical Software Engineering (ISESE'05)*, Noosa heads, Australia.
- [22] Mendes, E.: A Systematic Review of Web Engineering Research. Paper presented at the *ACM/IEEE International Symposium on Empirical Software Engineering*, Noosa heads, Australia. 2005.
- [23] Pedreira, O., Piattini, M., Luaces, M. R., y Brisaboa, N. R.: A systematic review of software process tailoring. *ACM SIGSOFT Software Engineering Notes*, vol. 32, pp. 1-6, 2007.
- [24] Petticrew, M. y Helen, R. *Systematic Reviews in the Social Sciences. A practical Guide*. Blackwell Publishing, UK, 2006.
- [25] Pfahl, D. A., Laitenberger, O. B., Ruhe, G. C., Dorsch, J. D., y Krivobokova, T. E.: Evaluating the learning effectiveness of using simulations in software project management education: Results from a twice replicated experiment. *Information and Software Technology*, vol 46, n° 2, pp: 127-147, 2004.
- [26] Sjöberg, D. I., Hannay, J., Hansen, O., Kampenes, V., Karahasanovi, A., Liborg, N., et al.: A Survey of Controlled Experiments in Software Engineering. *IEEE Transactions on Software Engineering*, vol 31, n° 9, pp: 733-753, 2005.
- [27] Staples, M., Niazi, M.: Systematic Review of Organizational Motivations for Adopting CMM-based SPI. *Information and Software Technology*. vol 50, n° 7-8, pp: 605-620, 2008.
- [28] Staples, M., y Niazi, M.: Experiences using systematic review guidelines. *Journal of Systems and Software*, vol. 80, n° 9, pp. 1425-1437. 2007.
- [29] Succi, G. A., Pedrycz, W. B., Djokic, S. B., Zuliani, P. A., y Russo, B. A.: An empirical exploration of the distributions of the Chidamber and Kemerer object-oriented metrics suite. *Empirical Software Engineering*, vol 10, n° 1, pp: 81-103, 2005.